ELSEVIER

International Journal of Industrial Ergonomics 26 (2000) 231–248

International Journal of Industrial Ergonomics

www.elsevier.nl/locate/ergon

An analysis of the costs and benefits of a system for FAA safety inspections

Philip A. Hastings[a,*], Michael Merriken[b], William B. Johnson[b]

[a] USWeb Corporation, 3300 Northeast Erpy, NESTE 8A, Atlanta, GA, USA
[b] Galaxy Scientific Corporation, Atlanta, GA, USA

Abstract

Utility analysis is a special case of cost-benefit analysis in which the targeted benefit is an improvement in work performance. Utility analysis is ordinarily used to estimate the dollar value of implementing one hiring procedure over a previous hiring procedure with a lower validity. Utility analysis can also be extended to situations in which workforce productivity changes as a result of an organizational intervention. In the present study utility methods were applied to an organizational change within the Federal Aviation Administration (FAA) that attempted to improve the workflow of aviation safety inspectors. The On-line Aviation Safety Inspection System (OASIS) allows FAA safety inspectors to log inspections using portable computers. This study estimates performance gains in dollars after the implementation of OASIS. The study found that aside from qualitative benefits such as better usability, the new system also saved labor time (about 19.2% of an inspector's workday). Conservative estimation procedures based on time data estimated by aviation safety inspectors indicated a net value derived from saved labor cost in excess of $16 million over the course of four years.

Relevance to industry

Many types of qualitative and quantitative data can support the effectiveness of performance improvement programs. However, much of this data is poorly understood by decision-makers. Utility analyses provide well-understood financial evaluations of human capital investments that are readily comparable to other business investments.

Keywords: Cost–benefit; Utility; Finance; Investment; Information systems

1. Introduction

Because of recent pressures to streamline operations within the Federal Aviation Administration (FAA), Flight Standards has instituted a program that facilitates the essential workflow of Aviation Safety Inspectors (ASI's) through computerized work tools. The previous system of reporting inspections involved writing or typing paper forms, submitting the forms for data entry by another person, and transferring the data to the national database. The redesigned system is called the On-line Aviation Safety Inspection System (OASIS). OASIS is used to gather inspection data, access reference materials, and provide guidance during

* Corresponding author. Tel.: + 1-770-451-6675; fax: + 1-404-249-6894.

E-mail address: phastings@usweb.com (P.A. Hastings).

0169-8141/00/$ - see front matter
PII: S0169-8141(99)00068-2

the inspection. The new system can be considered a human capital investment (Flamholtz, 1985) with projected benefits of timesaving and ease of use as compared with the previous inspection process.

The OASIS system was deployed to about 15% of the inspector workforce and will soon be implemented among the entire workforce. An investigation of the costs versus benefits of the system was considered an important question prior to the authorization of further funding. This paper attempts to answer that question by applying utility analysis (Boudreau, 1991; Schmidt et al., 1982) and usability analysis (Wixon and Ramey, 1996; Nielsen and Mack, 1994). The next section presents the formulation of utility and usability as applied in the present context.

2. Utility and usability analysis

When there are many options to consider during a decision-making task, it is useful to evaluate the options with a common metric. Techniques to aid decision-making can be as simple as a list of the factors to "keep in mind" during a decision, or as complex as calculated scores for each decision alternative by importance, quality, and cost. Two terms used to describe structured models for making decisions are *cost–benefit analysis* (CBA) (Flamholtz, 1985; Hull, 1980; Bierman et al., 1981; Karat, 1996; Remenyi et al., 1995) and *multiattribute utility analysis* (MUA) (Edwards, 1977; Huber, 1980; Keeny and Raiffa, 1976). This paper will use the term *cost–benefit analysis* when referring to any type of structured method for evaluating decision options. *Utility analysis* will refer to a specific type of cost–benefit analysis in which the decision concerns a human capital investment. This distinction is made for convenience of presentation and does not reflect a consensus among writers in this field.

Formalizing one's decision model has a number of tangible benefits. These are listed in Huber (1980), and repeated in Boudreau (1991):

- The model makes explicit a view of the decision situation, and helps to identify the inadequacies of the corresponding implicit, mental model.
- The attributes contained in models serve as reminders of the information needed for consideration of each alternative.
- The informational displays and mathematical models serve to organize external memories.
- The models allow the aggregation of large amounts of information in a prescribed and systematic manner.
- The models facilitate communication and support to be gained from constituencies.

CBA has become widely accepted among business and governmental organizations. Although CBA has definite limitations, especially in the nonstandard way that the payoff function is derived and calculated, its potential for making decisions more rational is comforting to those who must make the decisions. In situations in which large amounts of money are at stake, the presentation of a cost–benefit analysis is the preferred way to demonstrate the reasoning behind expenditures. Governmental budgeting agencies increasingly demand these quantitative, explicit justifications.

Utility analysis (UA) is a specific type of CBA in which the objective is a decision related to some investment in human capital. This type of investment could include hiring workers, training them, changing their workflow, etc. The next section discusses the formulation of utility as applied in the present study.

2.1. Formulation of utility with a time-based measure of performance

An organizational change can have many objectives, ranging from improved job satisfaction to better productivity of workers. When the objective is increased speed of certain essential job activities, criteria are measurable and quantifiable. The first step in assessing a workflow change designed to increase speed is to identify all of the target tasks that will be affected. The next step is to measure the time that it takes to complete targeted tasks before and after the intervention. Time can be measured in multiple ways. Some methods include sampled task times, archival analysis of the number of production units processed over some period of time, and time estimates made by the sample of workers (Mundel, 1978; Spencer, 1986).

Given that one has used a reasonable measure of time for the target job activities and sampled across situations or people, one can present the results of the analysis in terms of the proportion of time saved ($\Delta\bar{t}$) per unit time after the intervention. This is accomplished by subtracting the mean post-intervention time ($\bar{t}_x$) from the mean pre-intervention time ($\bar{t}_o$), and dividing this quantity by the mean pre-intervention time.

$$\Delta\bar{t} = ((\bar{t}_o - \bar{t}_x)/\bar{t}_o).$$

If one wishes to present the total timesaving of the intervention, one simply adds together the mean pre- and post-intervention time estimates for all tasks, and uses the same equation on the resulting totals. This type of analysis might appear simplistic, but for the purposes of deciding whether to embark on an intervention it can be a useful decision aid.

A more definitive analysis of the intervention would be to collect time estimates from a representative sample of incumbents and perform a *t*-test on the pre- and post-intervention sample of estimates. In this case the decision-maker has not only average percentage timesaving estimates, but also information about whether the organizational design was a statistically significant improvement. One could also construct confidence intervals about the mean difference of the pre- and post-time estimates, indicating the level of uncertainty due to unreliability in individual estimates (Hays, 1988; Maxwell and Delaney, 1990).

Florin-Thuma and Boudreau (1987) developed a method of applying the utility equations typically used in employee selection situations to the situation where incumbent employees were trained for increased performance. Building upon the utility equation presented in the Florin-Thuma and Boudreau (1987) study, and substituting the time element for their production unit: we get

$$\Delta U = f\,\Delta\bar{t}\,\bar{w}N - C$$

where

- ΔU is the one-year utility achieved by the intervention in dollars,
- $\Delta\bar{t}$ is the average expected proportion of time saved per unit of time,
- C is the cost of implementing the intervention,
- N is the number of employees affected by the intervention,
- f is the fraction of total time that is affected by the intervention,
- $\bar{w}$ is average annual salary of employees affected by the intervention,

The above equation can be rewritten to accommodate capital budgeting principles such as discounting (Boudreau, 1983; Cascio and Morris, 1990; Hirt and Block, 1993). Assume that the base time period for rate and salary calculations is one year. Further assume that the intervention will have effects beyond one year. A restatement of the above equation would then look like

$$\Delta U = \sum_{k=1}^{K} \frac{[f\,\Delta\bar{t}\,N_k\bar{w}_k] - C_k}{(1+i)^k} \tag{1}$$

for the years $1,\ldots,K$ (all variables as previously defined). The variable i represents the discount rate. The divisor is the discount factor, an adjustment for the fact that money invested in human capital is an opportunity cost that would be interest bearing if invested in another area. Components for variable costs and tax structures can be included as well (see Boudreau (1991) or Cascio and Morris (1990) for a more thorough discussion). However, it was assumed that no variable cost was associated with the present fixed contract, and tax rates are irrelevant for government expenditures; those components are left out of the utility equations here for the purpose of clarity.

Another method of estimating total proportion of timesaving for a certain organizational change is to ask workers to estimate this quantity directly. This average proportional timesaving estimate will be signified as P. The utility calculation looks very similar to Eq. (2), except that P replaces $f\,\Delta\bar{t}$.

$$\Delta U = \sum_{k=1}^{K} \frac{[PN_k\bar{w}_k] - C_k}{(1+i)^k}. \tag{2}$$

Two methods of calculating utility are presented in this study. Eq. (1) represents a job analytical approach to the problem that combines different estimates of discrete job tasks to arrive at a final utility estimate, and will be referred to as "analytic estimation." Eq. (2) represents a more global estimation approach and will be called "proportional estimation."

Timesaving may be only one of the projected benefits of an organizational change however. Other benefits could be better design of the workspace or the integration of essential job functions. The field of human-computer interaction (HCI) seeks to understand and apply the principles of human factors design to aid people who interact with computer technologies (Preece et al., 1994; Boehm, 1988; Karat, 1996). One method for evaluating these technologies is called usability analysis. *Usability* is an attitudinal/psychological measure of the overall satisfaction, ease of use, and ability to accomplish critical functions of software (Wixon and Ramey, 1996; Nielsen and Mack, 1994). One fairly direct method of usability analysis is called heuristic inspection, because certain heuristic criteria are used to rate the performance of the software (Baeker and Buxton, 1987; Malik, 1995). The present study focused on two heuristic criteria believed to contribute to end user satisfaction: general usability (including interface design, program responsiveness, facilitation of work objectives), and functionality (performance of specific work functions). Evaluations of usability are only made after system implementation. Conclusions based on the usability analysis are presented with the appropriate limitations.

2.2. Hypotheses

1. Total utility for the organizational redesign effort will be positive when projected over a four-year software cycle.
2. Average usability of OASIS will be higher than the midpoint of the scale (moderate), after adjusting for importance of individual factors contributing to usability.
3. Average functionality of individual applications in OASIS will be higher than the midpoint of the scale (moderate), after adjusting for importance of factors contributing to functionality for each application.

3. Method

The first version of the OASIS system was developed in the beginning of 1996, and was deployed during the months of June through August 1996. The schedule of deployment was as follows:

Task	Wave	Timeline
• OASIS deployment and training of inspectors	1	June–August 1996
• OASIS follow-up evaluation	1	September 1996
• Pre-OASIS evaluation	2	May 1997
• OASIS deployment and training	2	May–October 1997

The first OASIS inspectors are referred to as the Wave 1 group, and the pre-OASIS inspectors are referred to as the Wave 2 group. The research design was entirely post-treatment (ex post facto) in the Wave 1 group, with some comparison variables used to assess the equivalence of groups prior to treatment. This design is subject to some threats of internal validity (Cook and Campbell, 1979); these threats are mentioned in the next section. Evaluations were collected from the Wave 1 group approximately two months after they had received the training and were using the new system. Evaluations of the Wave 2 group occurred approximately nine months after the evaluation of the Wave 1 group.

The Wave 1 group was evaluated with a post-treatment measure that included some comparison variables (age, tenure, job grade, and computer experience), measures of time, and measures of usability. The Wave 2 group was evaluated nine months after the evaluation of the Wave 1 group due to budgeting delays in the US Congress as well as union issues, which prevented the identification of deployment sites for the second wave of OASIS until April 1997. Although there are nontrivial threats to the internal validity of the study (such as the time lag between waves), the research is believed to have consequential validity. Results could impact decisions made concerning the effectiveness of the system and its capacity to help improve the safety of air transportation. Those conclusions depend in part upon the presently available data. A significant effort was made to reduce contaminating effects of unmeasured variables, but as in most field studies, the ideal research design was not achieved. Limitations of the design are made explicit, and where possible conservative adjustments were made for the lack of randomization and other contributions to selection bias.

3.1. Sample characteristics

Eighty Aviation Safety Inspectors (ASIs) from the Federal Aviation Administration were selected to be initial evaluators (Wave 1) of the OASIS system. The eighty inspectors were stratified across nine FAA offices representing most regions of the United States. Six of the offices were Flight Standards District Offices (FSDOs) that are responsible for the domestic regulation of operations, avionics, and maintenance of commercial and private aircraft. Three of the offices were International Field Offices (IFOs) that are responsible for the regulation of foreign aircraft operators. Complete data are available for fifty-nine of the inspectors in the Wave 1 group.

The comparison group (Wave 2) consisted of inspectors in offices that were scheduled to receive the OASIS software during the second deployment, from May–October 1997. Evaluations of this group were conducted prior to the second deployment of OASIS. The second deployment of OASIS was significantly larger with a total of 630 inspectors and office personnel. All members of an office received the OASIS system that had completed requisite training to become a certified ASI. District offices are believed to be representative of the population of US district and international field offices.

Fortunately there are data available on base salary, age, and job grade for the entire population of nonsupervisory inspectors. Table 1 shows the descriptive statistics for both the sample and the population. Base salary differs from total salary in that total salary includes a differential for locality. The comparison of base salary presented here equates these differences for locality. Base salary is completely determined by job grade and job step within that grade. As Table 1 shows, the sample estimates are slightly higher than the population statistics. Part of the selection of candidates who could receive the OASIS laptops involved the union requirement that inspectors had completed a minimum of orientation training, which usually lasts six months at the start of the inspector's tenure. This requirement might have contributed to

Table 1
Comparison of Sample versus Population Statistics

	Mean			
	Population	Entire Sample	Wave 1	Wave 2
Job grade	13.0	13.2	13.6	13.1
Step in grade	4.1	4.4	5.0	4.2
Base salary	$56,503.0	$59,383.0	$64,174.0	$ 58,258.0
Age	50.0	49.3	49.9	49.1
	S.D.			
Job grade	0.7	0.8	0.6	0.8
Step in grade	2.7	2.5	2.0	2.6
Base salary	$9324.0	$9400.0	$7048.0	$9537.0
Age	8.5	8.5	7.8	8.7
	N			
Job grade	2331.0	374.0	74.0	300.0
Step in grade	2331.0	371.0	73.0	298.0
Base salary	2331.0	368.0	70.0	298.0
Age	2331.0	371.0	73.0	298.0

Table 2
Separate Variance t-Tests for Differences on Select Demographic Variables[a]

	Mean Wave 1	Mean Wave 2	t	df	p2-tailed
Age	49.9	49.1	0.70	119	0.488
Years using computers	8.1	8.8	− 0.89	47	0.378
Computer experience	3.4	3.5	− 0.43	52	0.669
Tenure	10.1	7.6	3.69	135	0.000[b]
Job grade	13.6	13.1	5.46	135	0.000[b]
Salary	$68593	$61414	6.36	123	0.000[b]

[a]Note. df are approximate based on separate variance t-test methods (Blalock, 1972).
[b]$p < 0.0001$.

the higher sample statistics. No comparison data were available for tenure.

The Wave 2 group was compared to the Wave 1 group on the demographic and experience variables of age, tenure, job grade, and years using computers. An important question is whether the two groups are significantly different on any of these variables. Table 2 shows the results of the t-tests and the estimates used in the tests. The next few sections discuss particular aspects of this analysis. Separate variance t-tests were used due to unequal sample sizes.

3.1.1. Age in years

Age approximated a normal distribution using a Kolmogorov–Smirnov test. As can be seen from the previous table, there was no significant difference in mean age between groups. When interpreting any statistic that does not reject a null hypothesis, one cannot assume that there is no difference between groups on the basis of a nonsignificant test. However one can reasonably state that the average age of inspectors was nearly equivalent, at about 50 yr.

3.1.2. Tenure, job grade, and salary

There was a significant difference in tenure between waves. Wave 1 inspectors had a higher tenure on average than Wave 2 inspectors by about 2.5 yr. Tenure was positively skewed and the Kolmogorov–Smirnov test of normality was rejected, $d = 0.139, p < 0.05$ using the Lilliefors probabilities, indicating a nonnormal distribution. Lilliefors probabilities are recommended in cases for which population statistics are estimated from the sample.

One would expect salary to be correlated with tenure and age, and in fact this was the case (see Section 4, Results and Table 8). Salary was also significantly different between waves (Table 2). Salaries are rarely normally distributed and the present sample was not an exception. The Kolmogorov–Smirnov test indicated a slightly negative skew, $d = 0.103, p < 0.05$. The salary variable more closely approximated a normal distribution than tenure.

Job grade is assigned to government employees and signifies position ranking. Job grade is one of three factors that determine salary, so it is no surprise that this variable was different between waves given the difference in tenure and salary. The Wave 1 group was higher ranking on average than the Wave 2 group, with a mean difference of about one-half of a job grade.

3.1.3. Computer experience

Two measures of computer experience were used in the present analysis. The first measure of experience was the number of years the respondent had used a computer. The second measure was a scale constructed from the average of three Likert self-ratings of computer knowledge and frequency of use. Most users rated themselves to be frequent users of computers and considered themselves as having a moderate proficiency with computers. Frequency of use was negatively skewed, indicating that a greater number of inspectors rated themselves as frequent users than infrequent users. Cronbach's alpha, a measure of the internal consistency of the scale, was adequate at.80. Table 2 shows t-tests between waves on both variables. Results suggest that there was no significant difference between waves on either of the computer background variables. There was a slightly higher point estimate for the number of years using computers in Wave 2, but only by a half a year.

3.2. Measurement method

All measures were conducted by survey. Demographic data were collected at the time of training for the Wave 1 group, and were a part of the mailed surveys for the Wave 2 group.

Surveys were conducted in the Wave 1 group during follow-up evaluation meetings about the OASIS system. These meetings lasted between two and three hours. During the first 45 min of the meeting the surveys were completed. After a 15 min break, inspectors gave some verbal feedback about the usability of the software based on their answers to the survey. About half of the inspectors who had used the OASIS system in each office usually attended the evaluation meetings. Reasons for not attending usually resulted from unexpected work deadlines or emergency inspection activity. Those inspectors who did not participate in the meeting were strongly encouraged to complete and mail back a survey to the researcher. Instructions given verbally in the evaluation meeting were presented in a cover letter to the inspectors completing the survey outside of the meeting.

Surveys of the Wave 2 group were conducted entirely by mail. A cover letter, a demographics questionnaire, and a time estimate survey was mailed to every new inspector scheduled to use the OASIS system in the second wave of deployment, a total of 428 people. The office computer specialist at each office was asked to coordinate the distribution of the surveys. This resulted in fairly high response rates. Of the 33 offices contacted, all except two returned some completed questionnaires, a 94% office response rate. Of the 428 surveys mailed, 315 were returned for a 74% response rate, and of the 376 inspectors from offices that returned at least some data there was an 84% response rate. In sum, high response rates lend confidence that the sample was representative of the total Wave 2 population.

3.3. Measurement domains

There are two domains of measurement for the present study. The first domain is in the area of *productivity*, and the second is in the area of *usability*.

The productivity domain targeted the amount of time it took inspectors to complete essential tasks related to their inspections. To measure timesaving, survey questions were developed that estimated average completion time for discrete job tasks before and after using the OASIS system. Identification of discrete tasks was accomplished by use of subject matter experts (inspectors) involved in the development of OASIS. Job tasks fell into three sub-domains: (1) accessing critical information (such as handbooks and regulations), (2) filling out forms by hand and entering data to log inspection activities, and (3) data entry issues (such as correcting data and entering the same data more than once). Items were created that asked the question: "On average, how much time (in minutes) does it take you to …" followed by a target job task. Based on the estimates both before and after using OASIS, timesaving scores were calculated.

Percentage of timesaving associated with OASIS was measured by asking the question: "On average, what is the percentage of timesaving associated with …" followed by a listing of high-level job activities. Inspectors were also asked to divide their time into four global categories, to allow the researcher to adjust estimates in the calculation of utility. The four categories were: (1) entering data into a computer, (2) completing required letters, forms and other paperwork, (3) looking up information in tables, handbooks, orders, regulations, etc., and (4) all other job activities.

In the domain of usability, three areas were identified: (1) overall software design, (2) specific functionality, and (3) hardware performance. A minimum of four items was created to evaluate each area. Items were rated using two scales. The first scale was the performance of the characteristic, rated on a 1–5 Likert scale. The second scale was the importance of the characteristic in relation to the inspector's job, also rated on a 1–5 Likert scale. The characteristics were rated on two scales because it was believed that some items would be less important than others to the end user. This second rating of importance allowed for the selection of items without giving undue influence to the unused or unimportant characteristics of the software.

3.3.1. Utility scales (time estimates)

All questions of time estimates asked the inspectors to judge the amount of time it took them ("on average") to complete discrete job tasks. The questions were divided into three major areas: access of information, inspection logging tasks, and data entry tasks. The questionnaires differed only in that Wave 1 inspectors were asked to make estimates both for the period *prior* to using OASIS and for the *current* period, while the Wave 2 inspectors were asked to estimate times for only their *current method* of completing tasks. Following are brief descriptions of each of the scales that measured time.

3.3.1.1. Time accessing critical information. This scale consisted of the sum of three items that asked inspectors to estimate how much time (in minutes) it took to complete job tasks related to the lookup of necessary information required by the job, such as Federal Aviation Regulations, Airworthiness Circulars, and Handbooks.

3.3.1.2. Time completing inspection forms and paperwork. This scale was the sum of three items that asked inspectors to estimate the amount of time (in minutes) it took to fill out inspection forms and paperwork, including the lookup of tabled codes for the forms.

3.3.1.3. Time entering/correcting data. This scale consisted of the sum of two items that asked inspectors to estimate the amount of time (in minutes) it took to correct data while logging a PTRS activity, or enter the same information more than once.

These time estimates allowed the researcher to calculate difference scores (in minutes) between the pre- and post-OASIS implementations. Difference scores could be calculated in the Wave 1 group only, since this group could estimate both pre- and post-OASIS times. Time difference variables were calculated as the simple difference between the time estimates with and without using OASIS.

An exploration of the data revealed that a few of the inspectors in the Wave 2 group made large estimates for the time it took to complete tasks. Since the data represented time estimates the researcher assumed that some people would normally take much more time than others would to complete tasks (in other words some outlying data points might be accurate estimates). Therefore rule for deletion was specified so that only the most extreme values (more than six times greater than the distance from the mean to the 75th percentile) were removed. This procedure contributed to conservative benefit estimates since difference scores were reduced slightly. Table 3 shows the descriptive statistics for the variables after being trimmed of extreme values.

One might make the argument that pre-OASIS estimates made by the Wave 1 group are prone to error, since the inspectors were asked to make the estimate for a period occurring in the past. For this reason the same quantity was measured in the Wave 2 group (that had never used OASIS), as a check on the precision of the estimates in Wave 1. Separate variance t-tests were performed to investigate the possibility that the Wave 1 inspectors were significantly different than the Wave 2 inspectors.

Table 3
Descriptive statistics for time estimates, trimmed sample

	Mean	S.D.	N	− 95% C.I.	+ 95% C.I.
Time accessing information (without OASIS)	41.2	33.6	320	37.48	44.87
Time accessing information (using OASIS)	21.5	22.8	37	13.99	29.11
Time accessing information (difference)	18.0	38.0	35	4.94	31.06
Time paper and PTRS (without)	21.6	18.5	322	19.55	23.60
Time paper and PTRS (using)	12.4	10.4	40	9.11	15.74
Time paper and PTRS (difference)	3.1	12.4	36	− 1.09	7.31
Time data entry (without)	8.3	8.5	317	7.41	9.28
Time data entry (using)	3.7	3.1	35	2.66	4.77
Time data entry (difference)	3.5	7.5	34	0.89	6.11

Table 4
Separate variance t-tests for time estimates, trimmed sample

	Mean Wave 1	Mean Wave 2	t	df	p 2-tailed
Time accessing information (without OASIS)	38.7	41.5	−0.52	51	0.607
Time inspection paperwork (without)	15.3	22.6	−2.97	68	0.004[a]
Time data entry (without)	7.1	8.5	−1.02	51	0.312

[a] $p < 0.01$.

The results of those t-tests are shown in Table 4 below. The only significantly different mean was the inspection paperwork estimate. The average time estimates for each scale are slightly greater in the Wave 2 group than in the Wave 1 group.

One should note that the relative pattern of means is the same in both groups, namely that accessing critical information takes the longest amount of time, while the time to enter and correct data takes the shortest amount of time. This similar pattern of means between groups appears even at the item level, an encouraging finding suggesting that both groups are estimating similar quantities.

An important question concerns the dependability or accuracy of time estimates by the inspectors. Although there is no objective comparison to actual work time, this question could be addressed somewhat by examining the internal consistency of specific items within domains, as opposed to the correlation of items across domains. In sum, indirect evidence for the dependability of the estimates indicates that there is differentiation between domains at the item level, that the scales are internally consistent, and that the difference scores correlate positively with ratings of overall timesaving.[1] In general it seems that the scales measuring time (aside from some outlying estimates) compare well between groups. There is some evidence to suggest that the groups differ on one estimate of time. The analyst realizes that one may never prove the equivalence of groups, only reject the hypothesis of equivalence. However, based on the fact that just one of three t-tests rejected the equivalence hypothesis, and that the pattern of means was highly similar, the analyst concluded that the groups were similar enough to proceed with the benefit analysis comparisons.

3.3.2. Utility scales (percentage estimates)

Percentage estimates are made using four scales: percent of time saved during paperwork activities, percent of time saved due to the portable computer, percent of time spent within specific categories of work, and the overall percent of time saved in all work tasks. Scales that measured percentages usually asked the inspector to consider the time saved by the current method of completing a job task as compared with the previous method. An exception was the item that asked inspectors to divide their time into specific categories and estimate percentages for the categories. To allow for the possibility that the new method resulted in time loss rather than savings, inspectors were instructed to use negative percentage estimates for time loss. The following paragraphs describe the scales used for the percentage estimates.

3.3.2.1. Percent time saved using OASIS per activity. This item asked inspectors to estimate what percentage of time was saved during an inspection activity due to OASIS.

3.3.2.2. Percent time saved per activity due to work redesign. This item asked inspectors to estimate what percentage of time was saved during an inspection activity due to having their necessary

[1] The average Cronbach's alpha was 0.83 in the time accessing information domain, 0.59 in the inspection paperwork domain, and 0.87 in the data entry domain. For time estimates without using OASIS the item relationships were stronger within domain, averaging 0.55 within versus 0.37 across domains. For time estimates using OASIS, the average within domain item correlation was 0.49 while the average across domain correlation was 0.19.

work tools contained on a single portable computer.

3.3.2.3. Division of work tasks into percentages of time. This item asked the inspectors to divide their *total work time* into percentages (that added to 100%) among the following four categories:

- entering/correcting data,
- completing forms, letters, and paperwork,
- accessing critical information (such as FAR's and handbooks),
- all other work activities

By forcing the division of time into the above categories, the researcher was able to translate the time estimates (in min) to estimates of total value. This was the key to calculating the total benefit (or cost) of the OASIS system. Both the Wave 1 and Wave 2 groups made percentage estimates. Estimates for this item were required to sum to 100%. If a respondent's percentages did not sum to 100%, the case was not used in the analysis. A total of 45 inspectors had missing data, and 15 inspectors' percentages did not sum to unity. These cases were not used for relevant calculations of benefit.

3.3.2.4. Percent of total work time saved by using OASIS. This estimate followed the division of work tasks item. It was assumed that this sequencing of items would prime respondents to think about total timesaving in relation to *all* job activities, not simply the percentage of savings associated with completion of inspection forms or related activities. The item asked inspectors to estimate what percentage of *total work time* is saved due to OASIS applications and the portable computer. Although the item could only be answered by Wave 1 respondents, it provided a check on the precision of the benefit calculation from the time estimates given by the entire sample.

Examination of the data revealed outlying estimates as in the previous scales. The same rule was used to identify extreme values as discussed earlier. Table 5 shows the descriptive statistics for the percentage scales after trimming extreme estimates. The large drop in sample size for the percentage category estimates resulted from the deletion of respondents whose percentage estimates did not sum to 100%.

Table 5
Descriptive statistics for percentage estimates, trimmed sample

	Mean %	SD.	N	-95% C.I.	$+95\%$ C.I.
Percent data entry	13.6	09.1	312	0.126	0.146
Percent paper and forms	23.1	12.9	312	0.216	0.245
Percent access info	18.9	11.4	312	0.176	0.201
Percent all other job activities	44.5	20.2	312	0.422	0.467
Percent saving from OASIS apps	36.7	27.4	44	0.284	0.451
Percent saving from laptop	45.4	30.0	49	0.368	0.540
Overall percent savings, total work time	19.2	18.4	51	0.140	0.244

The *t*-test for differences between groups was calculated on the trimmed sample, and there was no significant difference found on any scale. See Table 6 for the results. Although there was no significant difference, the point estimates are slightly lower for the Wave 2 group than for the Wave 1 group on the percent category estimates, which represents another tendency to underestimate the benefit of OASIS given the formulation of benefit and cost that was used.[2]

3.3.3. Usability: item and scale analysis

Items relating to the usability of OASIS could only be answered by the Wave 1 group. For every item the inspectors made two ratings: *performance*

[2] One can understand this intuitively given the difference in percentage estimates for "all other job activities" in each group. This percentage represents the amount of an inspector's time that *cannot* be affected by the OASIS system. Wave 2 estimates this quantity to be nearly 45% of their time. Wave 1 estimates this to be 42% of their time. Therefore, OASIS has the potential to affect about 58% of an inspector's day according to the Wave 1 group (100% − 42%), but only about 55% of an inspector's day according to the Wave 2 group. Because the Wave 2 sample size is so much larger than the Wave 1 sample size, the average estimate across groups will be closer to the Wave 2 estimate, resulting in a smaller "potential" benefit.

Table 6
Separate variance t-test of percentage estimates between waves[a]

	Mean Wave 1	Mean Wave 2	t	df	p 2-sided
Percent data entry	0.155	0.133	1.54	59.3	0.129
Percent paper and forms	0.238	0.229	0.47	61.1	0.643
Percent access info	0.187	0.189	− 0.08	55.9	0.938
Percent all other job activities	0.420	0.449	− 1.01	63.7	0.319

[a] $p < 0.05$

and *importance*. Importance represented the inspector's judgement of the necessity of the function or characteristic of the software in relation to the inspector's job. The importance rating allowed the researcher to create a rule for including items into the final scale. The reasoning behind this rule was that performance ratings of the "unimportant" characteristics would contribute undue error to the scale. For example, most inspectors claimed they had not used the calendar function because they preferred their personal appointment books. Since the function was rated unimportant by most inspectors it was not included in the scale. The rule for inclusion was that the mean of the importance rating had to be at least one standard deviation above the point that represented moderate importance (a value of 3 on the 5-point Likert scale). For example, an importance rating with a standard deviation of 0.9 and a mean of 4.0 would meet the rule for inclusion, because $4.0 - 0.9 > 3$. A more complex multiplicative weighting scheme was assumed to yield a scale with more error than the use of the simpler additive method discussed here, since any error in the importance ratings would be compounded if used as a coefficient on the performance items.

3.3.3.1. General usability. All items that targeted the overall design of the OASIS system met the rule for inclusion. The four items that made up the scale pertained to overall system speed, general look and feel of the program, the ability to meet business objectives, and the ability to help improve productivity. All of the importance items were higher on average than the performance items; this was probably due to the fact that the four items described positive ideals for software design. The fact that the performance items were rated differently from the importance items suggests that the performance items were being evaluated against these positive characteristics.

Table 7
Descriptive statistics for usability and functionality scales, trimmed sample

	N	Mean	SD	− 95% CI	+ 95% CI
Usability	55	3.5	0.8	3.3	3.7
Functionality	39	3.2	0.9	2.9	3.5

Reliability analysis of the scale yielded a Cronbach's alpha of 0.809. The Kolmogorov–Smirnov test on the continuous variable was rejected, $d = 0.172$, $p < 0.01$ according to the Lilliefors probabilities, signifying that usability has a slightly negative skew. Table 7 shows descriptive statistics for both general usability and functionality scales.

3.3.3.2. Functionality. Analysis of the items that measured functionality indicated that some functions were very important, and some functions were unimportant or unused. The scale that was created included only those items measuring functions deemed important by the inspectors, as described by the rule for inclusion discussed previously. Four items had importance scores that were much higher than the other items. These items corresponded to the following functions of OASIS: (1) the Field Kit, an application for the entry of inspection data, (2) the Transfer Utility, an application for uploading data to the mainframe system, (3) the Job Aid, a step-by-step checklist of job tasks, and (4) the Assistant, a standard form letter application for notifying operators about investigations. Cronbach's alpha was 0.81 indicating adequate reliability, and the Kolmogorov–Smirnov test of normality was nonsignificant, suggesting that the Functionality scale approximated a normal distribution.

4. Results

Displayed below are the correlations of the most important scales used in the analysis.

The correlational pattern among the major variables of the analysis is not surprising for the most part. One would expect tenure, salary, and age to correlate. The highest correlation among the three is that of salary and tenure at 0.75 (See Table 8). There is a positive relationship between years using computers and self-rated proficiency. Additionally, the higher a person rated usability, the higher their estimate of timesaving. The only peculiar correlation is the -0.41 relationship between years using a computer and functionality ratings. On the whole however, the variables appear to have a predictable correlational structure, suggesting that the scales are performing as intended.

4.1. Utility results

This section describes the method and results of deriving the total benefit value in dollars given the data collected. The first subsection details the equations used to calculate the benefit and cost. The next subsection details the impact that the scale and sample characteristics would have on interpretation of the estimates. The last subsection gives the tables of total and partial benefit estimates.

4.1.1. Calculation of benefit and cost

Now that the data have been explored for distributional assumptions and differences between groups, the calculation of a benefit value is fairly straightforward. Recall that the two formulations for utility developed in this paper are:

1. Analytic estimation:

$$\Delta U = \sum_{k=1}^{K} \frac{[f \Delta \bar{t} N_k \bar{w}_k] - C_k}{(1+i)^k},$$

2. Proportional estimation:

$$\Delta U = \sum_{k=1}^{K} \frac{[P N_k \bar{w}_k] - C_k}{(1+i)^k}.$$

These equations represent a simple concept. Essentially one finds the net present value of the investment by determining the total benefit in dollars for each year (the quantity in brackets), subtracting the total cost for each year (C_k), and summing over years. The divisor is simply the discount factor (an adjustment for the discount rate on future value). Two variables are open to interpretation by the analyst, $\Delta \bar{t}$ and f. The assumptions used in calculation of these values can lead to liberal or conservative estimates. It can be shown that the most reliable calculation of f is the sum of the sample estimates of percent of total work time devoted to activities that can be affected by the organizational change. Thus, $f = \text{OTE1A}_{12} + \text{OTE1B}_{12} + \text{OTE1C}_{12} = 0.555$ (Table 11). This is the fraction of total time that can be affected by the OASIS system. It can also be shown that the most conservative calculation of $\Delta \bar{t} = (\text{TAC}\Delta_1 + \text{TPT}\Delta_1 + \text{TDE}\Delta_1)/(\text{TACW}_{12} + \text{TPTW}_{12} + \text{TDE W}_{12}) = 0.346$. Conservatism is the preferred choice when estimating benefit or cost, since it is best to err on the side of caution.[3] Table 9 contains all of the means and sample sizes from the scales used in the calculations, excluding extremes.

[3] The reason this is the most conservative method is that the difference scores are calculated within person, but the pre-OASIS divisor (total time for activities before using the system) uses the whole sample estimate. The assumption is that difference scores must be dependent on the person making the estimates. Note that one could use group means to arrive at an "average" difference score, but this eliminates the dependency of the two quantities. Calculating the average difference from group averages risks the possibility that either group is under- or over-estimating value and could result in error. The divisor chosen for this method was selected because it had the largest value (making the final estimate more conservative) as well as the largest sample size (making it the most reliable estimate of true time. The analyst chose to select a large sample point estimate that would tend to underestimate benefit rather than use a more consistent value that might have the unfortunate consequence of overestimating benefit.

Table 8
Correlations of major variables

	Age	Tenure	Salary	Computer Exp.	Computer Years	Usability	Functionality	Δt Access Info	Δt Forms/ paperwork	Δt data entry	Overall timesave %
Age	1.00										
Tenure	0.45[b]	1.00									
Salary	0.36[b]	0.75[b]	1.00								
Computer Exp.	−0.27[b]	−0.20[b]	−0.07	1.00							
Computer Years	−0.06	−0.00	0.07	0.44[b]	1.00						
Usability	−0.15	−0.22	0.11	0.09	−0.11	1.00					
Functionality	0.21	0.15	0.26	0.02	−0.41[a]	0.74[b]	1.00				
Δt access info	0.14	0.13	0.24	−0.04	−0.34	0.62[b]	0.60	1.00			
Δt Forms/ paperwork	0.18	0.26	0.01	0.10	0.14	0.41[a]	0.18	0.40[a]	1.00		
Δt Data entry	−0.04	−0.15	−0.07	0.30	−0.04	0.50[b]	0.28	0.58[b]	0.46[a]	1.00	
Overall timesave %	−0.10	0.08	0.19	0.16	0.03	0.58[b]	0.48[b]	0.44[b]	0.52[b]	0.57[b]	1.00

[a] $p < 0.05$.
[b] $p < 0.001$.

4.1.2. Fixed and variable components for all utility equations

We may now construct the tables of fixed and variable values for the analysis, Tables 10 and 11. The cost data in are taken from the OASIS contract, Work Orders 1, 2, 3, and 5. All salary data were calculated based on the current sample and year. Year 0 is set to be the year 1997 (this adjusts net benefit to 1997 dollars), with the further simplifying assumption that all expenditures (including salary payments) are made in the beginning of the year. This assumption can be shown to underestimate actual benefit as long as the total salary cost is greater than total program cost, since wages distributed throughout the year would gain interest prior to payment (a benefit that is not added to the estimated value). Utility is calculated for a 3-year period after the initial investment.

4.1.3. Estimates of benefit and cost

With the values from the previous section we may now calculate the utility estimates. Table 12 details the calculation of utility for the analytical method of estimation as described by Eq. (1), and Table 13 details the calculation for the proportional method as described by Eq. (2).

Hypothesis 1 predicted that the utility estimate of net benefit would be positive when calculated over a three-year software lifecycle. As Table 12 demonstrates, the analytic estimation procedure predicts 16.4 million dollars present value. Remarkably, the proportional estimation procedure (Table 13) closely corresponds to the analytic estimation procedure value of 16.4 million dollars. The parameters that determine value in each method are separate estimates of very different quantities made by inspectors. The proportional estimate uses as a parameter the mean response to a single global timesaving item. The analytic estimate combines change scores across essential job domains and multiplies by the fraction of time that can potentially be affected by the system.

A final note about the estimated benefit. With the data above, one may easily solve for the "break-even" scenario in which the investment perfectly equals the return. One simply works backward from the total value by setting it at zero and constraining the estimates of $\Delta\bar{t}$ to be equal across

Table 9
Means for f and $\Delta\bar{t}$ calculations, trimmed sample

Estimate	Code	Wave 1	N	Wave 2	N	Waves 1,2	N
Time access info (without OASIS)	TACW	38.650	40	41.535	280	41.174	320
Time access info (using OASIS)	TACU	21.514	37			21.514	37
Time access info difference	TACΔ	18.000	35			18.000	35
Time paper and PTRS (without)	TPTW	15.273	44	22.570	278	21.573	322
Time paper and PTRS (using)	TPTU	12.425	40			12.425	40
Time paper and PTRS Difference	TPTΔ	3.111	36			3.111	36
Time data entry (without)	TDEW	7.103	39	8.515	278	8.341	317
Time data entry (using)	TDEU	3.714	35			3.714	35
Time data entry difference	TDEΔ	3.500	34			3.500	34
Percent saving from OASIS	TPSI	0.367	44			0.367	44
Percent saving from laptop	TPS2	0.454	49			0.454	49
Percent data entry	OTE1A	0.155	43	0.133	269	0.136	312
Percent paper and forms	OTE1B	0.238	43	0.229	269	0.231	312
Percent access info	OTE1C	0.187	43	0.189	269	0.188	312
Percent all other job activities	OTE1D	0.420	43	0.449	269	0.445	312
Overall percent savings, total work time	OTE2	0.192	51			0.192	51

Table 10
Fixed values for the cost/benefit analysis

	Formula/definition	Calculation/source	Value(s)
$N_{k=-1}$	Number of employees affected by the investment	Government contract	80
$N_{k=0,1,2}$	Number of employees affected by the investment	Government contract	710
$C_{k=-1}$	Cost of investment year 1	Government contract, billed value	\$2 128 915
$C_{k=0}$	Cost of investment year 2	Government contract, funded value	\$7 610 432
i	Discount rate	Federal Reserve Bank	5%

Table 11
Variable values for the cost/benefit analysis

	Formula/definition	Calculation /source	Value(s)	SD	N
$\Delta\bar{t}$	Estimate of proportion of timesaving/unit time	$(\mathrm{TAC}\Delta_1 + \mathrm{TPT}\Delta_1 + \mathrm{TDE}\Delta_1)/(\mathrm{TACW}_{12} + \mathrm{TPTW}_{12} + \mathrm{TDEW}_{12})$	0.346	0.38	35/312
f	Fraction of total time that can be affected by the investment	$\mathrm{OTE1A}_{12} + \mathrm{OTE1B}_{12} + \mathrm{OTE1C}_{12}$	0.555	0.202	312
$\bar{w}_{k=-1}$	Average annual wage year 1	Federal worker salary schedule	\$60969	\$9775	368
$\bar{w}_{k=0}$[a]	Average annual wage year 2	Federal worker salary schedule	\$62780	\$10074	368
$\bar{w}_{k=1}$[a]	Average annual wage year 3	Federal worker salary schedule	\$64645	\$10373	368
$\bar{w}_{k=2}$[a]	Average annual wage year 4	Federal worker salary schedule	\$66565	\$10682	368
$\mathrm{SD}y_{k=-1}$	SD job performance in dollars	0.4 $(\bar{w}_k)$	\$24366	\$3910	368
$\mathrm{SD}y_{k=0}$	SD job performance in dollars	0.4 $(\bar{w}_k)$	\$25112	\$4030	368
$\mathrm{SD}y_{k=1}$	SD job performance in dollars	0.4 $(\bar{w}_k)$	\$25858	\$4149	368
$\mathrm{SD}y_{k=2}$	SD job performance in dollars	0.4 $(\bar{w}_k)$	\$26626	\$4273	368

[a]Projected wages were based on a 2.97% per year increase (Federal Office of Compensation, 1997).

Table 12
Analytical estimation of yearly and total net value for OASIS through 1999 by

$$\Delta U = \sum_{k=1}^{K} \frac{[f \Delta \bar{t} N_k \bar{w}_k] - C_k}{(1+i)^k} \tag{1'}$$

	Conservative Calculation of $\Delta\bar{t}$								
Year	k	f	$\Delta\bar{t}$	N_k	$\bar{w}_k$	$[(1+i)^k]^{-1}$	Benefit	C_k	Net value
1996	−1	0.555	0.346	80	\$60969	1.050	\$983462	\$2128915	−\$1145453
1997	0	0.555	0.346	710	\$62780	1.000	\$8559507	\$7610432	\$949075
1998	1	0.555	0.346	710	\$64645	0.952	\$8393366	\$0	\$8393366
1999	2	0.555	0.346	710	\$66565	0.907	\$8231532	\$0	\$8231532
								Total benefit	\$16428519

Table 13
Proportional estimation of yearly and total net value for OASIS through 1999 by

$$\Delta U = \sum_{k=1}^{K} \frac{[P N_k \bar{w}_k] - C_k}{(1+i)^k} \tag{2'}$$

Year	k	P	N_k	$\bar{w}_k$	$[(1+i)^k]^{-1}$	Benefit	C_k	Net value
1996	−1	0.192	80	\$60969	1.050	\$983247	\$2128915	−\$1145668
1997	0	0.192	710	\$62780	1.000	\$8565162	\$7610432	\$954730
1998	1	0.192	710	\$64645	0.952	\$8399569	\$0	\$8399569
1999	2	0.192	710	\$66565	0.907	\$8237178	\$0	\$8237178
							Total benefit	\$16445809

years. The break-even point for $\Delta\bar{t}$ is 0.129. This means that if inspectors save only about 13% of their time completing forms and paperwork, accessing critical information, and entering data (or 7% of their total work time), the investment would have been exactly worth the cost.

4.2. Usability results

Hypotheses 2 and 3 both predicted that the average ratings of the usability and functionality of OASIS would be higher than the moderate midpoint of the scale. Table 7 shows the 95% confidence intervals (CIs) of the means for both of these variables. The confidence interval about the mean specifies a region above and below the point estimate where it would be expected that 95% of the sample mean estimates would fall (given unlimited sampling). The CI for usability did not include the moderate midpoint of the scale – in other words, we can be 95% confident that the mean rating of usability was above average. The same cannot be said for the functionality rating. Though the mean estimate was above 3.0, the 95% confidence interval included this midpoint. Given the point estimate and the standard error about the mean, we can say only that we are about 89% confident that the mean rating of functionality was above average.

5. Discussion

Hypothesis 1 predicted that the benefit of the OASIS system would outweigh the cost, and even with multiple estimation procedures the net benefit was positive. Hypotheses 2 and 3 predicted that in terms of usability and functionality the inspectors would favorably receive the OASIS system. These hypotheses were supported by the data.

The estimate of overall timesaving was a single, global item on the survey: "What percent of your total work time is saved by using OASIS applications and the portable computer?" Respondents were primed to think about *total* work time by the prior question on the survey that asked inspectors to divide their entire work time into specific categories. This ordering of questions was designed to reduce the tendency to overestimate timesaving by getting respondents to think about all work tasks, not just those that pertained to the new system. The mean for that question was lower than means for other questions about timesaving, suggesting that the priming technique might have been successful. Perhaps this priming method could be used in future studies of timesaving to determine whether asking workers to divide their total time into work categories helps to stabilize more global items that follow. More research is needed in this area.

Inspectors estimated that they saved about 19.2% of their total time due to the OASIS system. The variation in this estimate was large, and there were negative estimates by some users. This produced a wide confidence interval for the estimate, with a lower boundary of 14.2% timesaving and an upper boundary of 24.2% timesaving. Even at the lowest boundary of timesaving however, the benefit was slightly more than \$1 million present value. This corresponds to a moderate return on investment of 10.3%, calculated by dividing the net value of the benefit by the initial cost.

Furthermore, recall that the $\Delta\bar{t}$ estimate represented the proportion of timesaving that occurred in relation to work activities affected by OASIS, not total work timesaving. One can translate this partial estimate to a total estimate by simply multiplying by the fraction of total time that *could* be affected by OASIS (f in Table 11). Using the estimate of $\Delta\bar{t} = 0.346$, and multiplying this by $f = 0.555$, yields an estimate of total timesaving $= 0.192$, or 19.2%. The fact that two estimates of the same quantity – one calculated from difficult questions about work time and the other derived from a single item – converge to the tenth of a percent is a convenient coincidence, but one that lends a measure of credibility to the time estimates.

Utility analysis makes a number of assumptions about the nature of work, the reliability of point estimates, and the projected benefits of work redesign. This study attempted to take a conservative view of the data, and made explicit the decisions that were made concerning the estimation process. The study was conducted entirely by survey and was subject to threats of internal validity (such as time and cohort effects) that cannot be discounted. The researcher attempted to minimize these problems in selection by the use of conservative downward adjustments on the estimates of utility. Sample characteristics that contributed to the tendency to underestimate benefit were the higher estimates of pre-OASIS task time by the larger Wave 2 group (as opposed to the slightly more conservative estimates made by the Wave 1 group). Because of this fact, difference scores were divided by a larger value, resulting in a smaller size of effect.[4] Additionally, the estimates of work time percentages (divided into categories) favored a downward bias, since the Wave 1 estimates of work time that could be affected by OASIS (f) was greater than the estimates made by Wave.[2] Other measurement assumptions were made in order to calculate the monetary value of the new system. These assumptions generally impacted the estimates in a conservative manner.[5]

[4] See footnote 3.

[5] Some of the critical assumptions were as follows. *The value of human resources can be approximated by average salary.* Many texts propose that salary significantly underestimates the value of human resources, since salary does not include other expenditures such as benefits and overhead costs. These texts often recommend multiplying salary by a factor at least twice salary. The analyst decided not to use a multiplier in order to remain as conservative as possible. (Flamholtz, 1985; Spencer, 1986; Phillips, 1983; Frank, 1984). *Workers who save time by using a new method to accomplish work tasks will use the saved time to do productive work.* To the extent that timesaving is not utilized for other work tasks, the value of the system will be lessened. *Sampling worker's perceptions of task times will approximate true task times.* Timed observations of work tasks are known to be unreliable when used for the analysis of white collar or managerial jobs (Cascio, 1991) due to the number of cognitive tasks. Instead of actually measuring the amount of time it takes to perform work tasks, this study relied on the estimates given by the people that completed the work.

6. Conclusion

Viewing the analysis from a business perspective, initial outlay should yield an annual return on investment of 42% over four years. Finding a benefit of such a large amount may cause the belief that the estimate cannot be accurate. However, when one looks at the total salary expenditure across the inspector population (more than \$150 million per year), a benefit of \$16 million over the course of four years does not seem so unrealistic. One should also keep in mind that the estimates simply represent *potential* benefit due to saved labor cost. Should inspectors not save time in during work, or not convert their saved time to productive work, the utility of the system is decreased. More research will be needed in relation to actual units of production to determine the true benefit of the new system.

Inspectors estimated that the *total* amount of work time saved by OASIS was 19.2%, which corresponds to more than $1\frac{1}{2}$ hours per day, 13 hours per week, or 17 weeks per year. One should also note that benefit is calculated here as the percent of total timesaving on the job. This value does not include intangible benefits related to quality of work life, ability to access current information quickly, and potentially improved safety.

Inspectors who used OASIS during the first year of deployment rated the system moderately well on a number of dimensions. Comments during the follow-up evaluations seemed to confirm this generally positive evaluation. Inspectors especially valued the ability to carry work tools while on trips in the field, and also their ability to use the reference disc while away from the office. The surveys also contained space for comments and an informal perusal of these revealed a mostly positive response to the new system. A relatively modest effort to improve the design of the inspector's work tools seemed to result in a fairly dramatic reduction on the amount of time to complete critical duties. This conclusion is supported by a number of factors, including the comments made by the inspectors themselves. Based on the available data, the benefit of OASIS appears to outweigh the cost.

Acknowledgements

The first author wishes to thank Dr. Bart Osburn for his helpful guidance. Portions of this manuscript were excerpted from the first author's dissertation.

References

Baeker, R.M., Buxton, W.A.S., 1987. Readings in Human-Computer Interaction. Morgan Kaufman, San Mateo.

Bierman Jr., H., Bonini, C.P., Hausman, W.H., 1981. Quantitative Analysis for Business Decisions. Irwin, Homewood.

Boehm, B., 1988. The spiral model of software development and enhancement. IEEE Computer 21, 61–72.

Boudreau, J.W., 1991. Utility analysis for decision in human resource management. In: Dunnette, M.D., Hough, L.M. (Eds.), Handbook of Industrial and Organizational Psychology, Vol. 2. Consulting Psychologists, Palo Alto, pp. 621–745.

Boudreau, J.W., 1983. Economic consideration in estimating the utility of human resource productivity programs. Personnel Psychology 36, 551–576.

Cascio, W.F., 1991. Applied Psychology in Personnel Management. Prentice-Hall, Englewood Cliffs, NJ.

Cascio, W.F., Morris, J.C., 1990. A critical reanalysis of Hunter, Schmidt, and Coggin's (1988) "Problems and pitfalls in using capital budgeting and financial accounting techniques in assessing the utility of personnel programs." . Journal of Applied Psychology 75, 410–417.

Cook, T.D., Campbell, D.T., 1979. Quasi-Experimentation: Design and Analysis Issues for Field Settings. Houghton Mifflin, Boston.

Flamholtz, E.G., 1985. Human Resource Accounting 2nd Edition. Jossey-Bass, San Francisco.

Florin-Thuma, B.C., Boudreau, J.W., 1987. Performance feedback utility in a small organization effects on organizational outcomes and managerial decision processes. Personnel Psychology 40, 693–713.

Hays, W.L., 1988. Statistics 4th Edition. Holt, Rinehart, and Winston, Fort Worth.

Hirt, G.A., Block, S.B., 1993. Fundamentals of Investment Management 4th Edition. Irwin, Boston.

Huber, G.P., 1980. Managerial Decision Making. Scott Foresman, Glenview.

Hull, J.C., 1980. The Evaluation of Risk in Business Investment. Pergamon Press, Oxford.

Karat, C.-M., 1996. Cost–benefit analysis of usability engineering. Paper presented at: Usability Engineering: Industry-Government Collaboration for System Effectiveness and Efficiency (February 26, 1996). Sponsored by the National Institutes of Standards and Technology.

Keeny, R.L., Raiffa, H., 1976. Decisions with Multiple Objectives: Preferences and Value Tradeoffs. Wiley, New York.

Malik, N.A., 1995. Put your interface to the test: what makes a good interface "good"? VB Tech Journal.

Maxwell, S.E., Delaney, H.D., 1990. Designing Experiments and Analyzing Data: A Model Comparison Perspective. Wadsworth, Belmont.

Mundel, M.E., 1978. Motion and Time Study: Improving Productivity, 5th Edition. Prentice-Hall, Englewood Cliffs, NJ.

Nielsen, J., Mack, R., 1994. Usability Inspection Methods. Wiley, New York.

Phillips, J.J., 1993. Handbook of Training Evaluation and Measurement Methods. Gulf Publishing, Houston.

Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., Carey, T., 1994. Human-Computer Interaction. Addison-Wesley, Reading, MA.

Remenyi, D., Money, A., Twite, A., 1995. The Effective Measurement and Management of IT Costs and Benefits. Butterworth-Heinemann, Boston.

Schmidt, F.L., Hunter, J.E., Pearlman, K., 1982. Assessing the economic impact of personnel programs on work-force productivity. Personnel Psychology 35, 333–348.

Spencer, L.M., 1986. Calculating Human Resource Costs and Benefits. Wiley, New York.

Wixon, D., Ramey, J., 1996. Field Methods Casebook for Software Design. Wiley, New York.